

AN ANALYSIS OF WEIGHT GAIN AMONG CALL CENTER WORKERS

Statistical Analysis Plan

Elizabeth Anderson, Gokcen Buyukbas, Xiao Liu

Introduction

Obesity-related diseases are costly for businesses, incentivizing interventions to reduce obesity. We examine factors related to weight gain in one specific case of employees working at one call center. First, we will examine multiple measures of weight gain, in addition to explanatory variables, such as the time of shift and exercise levels. To identify associations between factors, such as exercise and shift time, and weight gain, we use random forest regression and full information maximum likelihood estimation.

Analysis Population

Previous research identified female call center employees as being particularly at risk for musculoskeletal symptoms compared to employees working in other occupations. The present study surveyed employees at a call center in the Southeastern United States. Missing data were relatively common in the dataset, and after dropping respondents who had missing data on all values of interest ($N=40$), we were left with a sample of 352 respondents. Descriptive statistics of the sample can be found in Table 1.

Table 1: Baseline Covariates

| Covariate | Notes | Hypothesis |
|------------------------|---|---|
| Shift | The time when the subject starts to work every day | subject who has an early schedule is more likely to gain weight |
| Age | Age of subject | the older the subject is, the more likely the subject gain weight(since older people tend to not do exercise) |
| Height | Height of subject | height may not have a huge impact on weight gain |
| Pounds Gained | How much weight has the subject gained since starting the job (potential response?) | |
| BMI | Body mass index (potential response) | |
| Vigorous Exercise Time | Vigorous MET-minutes used per week | the longer time subject spends on doing vigorous activities per week, the less likely the subject gain weight |

| | | |
|------------------------|---|---|
| Moderate Exercise Time | Moderate MET-minutes used per week | the longer time subject spends on doing moderate activities per week, the less likely the subject gain weight |
| Walking Exercise Time | Walk MET-minutes used per week | the amount of time subject spends to walk may not have huge impact on weight gain |
| Total Metabolic Time | total of vigorous, moderate and walking MET-minutes used per week | the longer time subject spends on doing exercises per week, the less likely the subject gain weight |
| Weight Gain | if the subject had any weight gain since starting job(yes/no) | |
| Gender | gender of the subject (Male/Female) | gender of the subject may have very little effects on weight gain |

Specific Aims

Specific Aim #1: Summarize the descriptive differences between weight gain and demographic and employment measures

- 1.1 Examine the distributions of two potential dependent variables (a binary measure of weight gain over the survey period and a linear measure of pounds gained over the study period).
- 1.2 Compare the population of workers who gained weight compared to workers who did not gain weight across social and employment characteristics of interest. For categorical variables (shift time and gender), we will consider Pearson's χ^2 tests for significance. For continuous variables (total MET exercise time and age), we will use t-tests for significant differences.
- 1.3 Produce histograms for age and MET exercise time to examine the distributions of the continuous predictors. We will also produce histograms comparing group differences by weight gain.
- 1.4 Investigate missingness in the dataset through regression models predicting missingness on at least one of the predictor variables.

Specific Aim #2: Identify individual level factors associated with weight gain during the study period. We will use two regression methods capable of adjusting for missing data .

- 2.1 Using the dependent variable selected in SA2, we will use full information maximum likelihood estimation to examine the direction and magnitude of the association between the independent variables and a measure of weight gain. We may consider transformations of the variables if needed. We will also use random forest to analyze the data, and see if the conclusions from random forest match the conclusions from the FIML model.

Conclusion

In general, we want to find what variables affect the chance of gaining weight and how these variables affect the weight gain. We will select appropriate variables as responses first, then produce plots to illustrate the relationships between response and dependent variables that are listed in the Baseline Covariate Table. We will use the knowledge obtained from graphs to select appropriate predictors, and apply full information maximum likelihood regression to examine the relationship between responses and predictors. We will also apply random forest to analyze the same data set, and see if these two models give us consistent results.